Statistics Assignment :

True (a option)

Central limit theorem (a option)

Modeling event time data (a option)

The square of a standard normal random variable follows what is called chi-squared distribution (c option)

Poisson (c option)

False (b option)

Hypothesis (b option)

0 (a option)

Outliers cannot conform to the regression relationship (c option)

BRIEF ANSWERS :

10) The normal distribution is a bell shaped curve data distribution . Possessing a standard deviation of one. The data is supposed to be symmetric and such that the mean, mode as well median is same .

It is nothing but the gaussian distribution , the skewness in the distribution also predicts the outliers and one need to drop it to have best accurate model .

11) Talking about missing data , we need to check whether the data is classical or not , had the data been some random numbers , being independent we can simply fill the data by mean of values of the

columns or the median or mode, having knowledge which would best fit. Next one can just drop the missing data as and when required so as to prevent the overfitting of model.
pandas.fillna(column[name].mean())

pandas.dropna()

These are the some suggested method .

12) We take two versions of an app. , just to check which one would best fit means which of the two versions is better . In this just the interface is modified, that change could be a complete new visual of the web page or some modifications just, then the interface is exposed to users (visitors). Half of the visitors are shown the old version and half the new modified version, then based on statistical tools the data is taken weather there is some positive effect or not or may be negative, and hence the best version is set up final so as to bring (increase)

Business opportunities.

13) The mean of the values is at times not suggested and used because, it robs the relationship between that variable and other variable, that is correlation , which at times leads to overfitting as well. Also the outlier data points will have a significant impact on the mean . Hence he model accuracy will be high but the interpreted data will be wrong .

14) The process of finding the relationship between dependent and independent variable is called the linear regression , both the variables should be quantitate , and they are checked by fitting in a straight line ,

The data is taken on x axis and y axis and the plotted points are fitted (best) in a line that predicts the future data and dependency. The variation from the actual data is measured and its square is taken to check the deviation.

15) There are 2 major branches of statistics :

Descriptive Statistics: It involves the collection and presentation of the data. This involves the choosing of model and making the data ready to predict , that is avoiding the missing values , dropping the parameters that wont help the least .

Inferential Statistics: It involves the prediction of the future data that best fits, in this part one draws the conclusion , the generalization of the data by taking small samples also comes under this statistics branch.

THANK YOU